Received: 15 December 2020 | Revised: 30 April 2021 | Accepted: 12 June 2021

DOI: 10.1111/mila.12392

SUBMITTED ARTICLE

WILEY

Signalling, commitment, and strategic absurdities

Daniel Williams[1,2]

[1]Corpus Christi College, University of Cambridge, Cambridge, UK

[2]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK

Correspondence
Daniel Williams, Corpus Christi College, University of Cambridge, Trumpington Street, Cambridge CB2 1RH, UK.
Email: dw473@cam.ac.uk

Funding information
This work was supported by an Early Career Research Fellowship from Corpus Christi College, University of Cambridge.

Why do well-functioning psychological systems sometimes give rise to absurd beliefs that are radically misaligned with reality? Drawing on signalling theory, I develop and explore the hypothesis that groups often embrace beliefs that are viewed as absurd by outsiders as a means of signalling ingroup commitment. I clarify the game-theoretic and psychological underpinnings of this hypothesis, I contrast it with similar proposals about the signalling functions of beliefs, and I motivate several psychological and sociological predictions that could be used to distinguish it from alternative explanations of irrational group beliefs.

KEYWORDS
absurd belief, irrationality, motivated reasoning, self-deception, signalling, socially adaptive belief

> "Apparently irrational cultural beliefs are quite remarkable: they do not appear irrational by slightly departing from common sense, or timidly going beyond what the evidence allows. They appear, rather, like down-right provocations against common sense rationality" (Sperber, 1985, p. 85).

1 | INTRODUCTION

Here are some beliefs that—as of late 2020, at least—appear to be held by large numbers of people:

- Donald Trump is involved in a covert battle against a cabal of Satan-worshipping, cannibalistic paedophiles that run the Democratic Party.[1]
- Nineteenth-century American prophet Joseph Smith received a revelation that Christ's second coming will be followed by a millennium-long reign over Earth from Jerusalem and Missouri.[2]
- The Earth is a flat disc surrounded by a 150-ft wall of ice guarded by NASA employees.[3]
- It is possible to drink the blood of a man who is both God and the son of God and whose mother was a virgin.

To most people who do not hold such beliefs, they seem absurd. It is not that they appear mildly or even moderately irrational. In many cases, it is easy to understand how people form irrational beliefs. We deceive ourselves, succumb to wishful thinking, rely on unreliable heuristics, and so on. The beliefs just listed, however, appear *extremely* irrational. It is difficult to understand how people could seriously entertain such ideas, let alone endorse them. In these respects, there is a much-remarked-upon similarity between such beliefs and certain examples of clinical delusions of the sort that you find in conditions such as schizophrenia (McKay & Ross, 2020). Unlike clinical delusions, however, they appear to be embraced by many well-functioning individuals not suffering from any psychiatric or neurological disorder.

This gives rise to a classic puzzle: How could—and why do—seemingly well-functioning systems of belief formation give rise to absurd beliefs? It is widely held that the function of beliefs is to represent the world accurately (Fodor, 1975; Williams, 1973). Thus, it is plausible to think that our systems of belief formation evolved to be reliable—to be responsive to evidence and reasons in a way that is conducive to truth (Millikan, 1984). Of course, there is always a trade-off between efficiency and reliability and an individual's motivations can distort belief formation through wishful thinking and self-deception. Nevertheless, it is difficult to see how *extremely* irrational beliefs could be produced by functional systems of belief formation, even taking into consideration such distorting factors.

My aim in this article is to explore a partial solution to this puzzle that models at least some absurd beliefs as part of a signalling strategy used by cooperative groups to identify committed group members. According to this proposal—what I will call the "strategic absurdity hypothesis"—it is precisely because absurd group beliefs are viewed as absurd by outgroup members that their sincere and conspicuous endorsement functions as a credible display of ingroup commitment.

This strategic absurdity hypothesis is increasingly influential (Boyer, 2018; Kurzban & Christner, 2011; Mercier, 2020; Petersen, 2020; Petersen, Osmundsen & Tooby, 2020; Simler & Hanson, 2017; Williams, 2021) and fits with a growing trend in psychology and philosophy focusing on the expressive or signalling functions of beliefs more generally (Funkhouser, 2017, 2020; Kahan, 2013; McKay & Ross, 2020; Simler & Hanson, 2017; Williams, 2018). Nevertheless, it generates numerous questions that deserve more attention in the psychological and philosophical literature. How should it be understood, psychologically? In what sense, if any, do strategic absurdities qualify as genuine beliefs? How does this signalling strategy relate to and differ from other prominent signalling explanations of irrational beliefs? What determines the contents of strategic absurdities? Perhaps most importantly, what kind of evidence could be used to

[1]From Doward (2020).

[2]From the *Book of Mormon*.

[3]Reported from the documentary *Behind the curve*.

distinguish between this hypothesis and other more influential explanations of irrational group beliefs?

I take up these questions in this article. Section 2 introduces and clarifies the strategic absurdity hypothesis. Section 3 argues that strategic absurdities should be understood in terms of independently well-established processes of motivated cognition. Section 4 then motivates seven predictions that—at least if observed in conjunction with each other—would lend significant support to the strategic absurdity hypothesis as an explanation for a specific class of beliefs.

First, however, a brief note on absurd beliefs, the explanatory target of this article. I intend "absurd belief" as a term of art. As a working definition, such beliefs should be understood as *extremely irrational beliefs produced by functional systems of belief formation*. By "irrational" I mean epistemically irrational, a phenomenon which I will understand broadly in terms of information sampling and processing that systematically obstructs the acquisition of knowledge (Williams, 2020). Epistemic irrationality thus exists on a continuum. Extreme irrationality is characterised by striking deviations not just from optimal but from ordinary forms of inference and reasoning. By "functional" systems of belief formation, I mean systems that are either operating in accordance with their evolved function or in a way that is appropriately sensitive to the believer's motives and interests.

2 | THE STRATEGIC ABSURDITY HYPOTHESIS

In this section, I describe the characteristics of coalitions that incentivise signals of ingroup commitment (Section 2.1), I clarify the game-theoretic logic that underlies strategic absurdities (Section 2.2), and I introduce the strategic absurdity hypothesis itself and contrast it with alternative signalling explanations of irrational group beliefs (Section 2.3).

2.1 | Coalitions, cooperation, communication

Humans are a coalitional species (Boyer, 2018; Tooby & Cosmides, 2010). From small-scale shifting alliances and factions to larger-scale, more stable bands, tribes, sects, cults, religions, political parties, online communities, and so on, human social life is characterised by complex forms of intragroup cooperation and intergroup competition, and it has been for a long time. As a consequence, we appear to have inherited a powerful and multifaceted coalitional or "tribal" psychology (Pietraszewski, 2016; Richerson & Boyd, 2008; Tooby & Cosmides, 2010). For example, people automatically encode information about alliances in social interactions, they instinctively interpret benefits and costs to their coalition as benefits and costs to themselves, and they exhibit various forms of ingroup bias, including not just ingroup favouritism but also greater empathy for and trust of ingroup members, and greater distress by intragroup disagreement than intergroup disagreement (see Boyer, 2018, pp. 38–45, for a review).

Coalitions vary along numerous dimensions, including size, structure, cohesiveness, longevity, membership criteria, and so on, but all share in common the use of within-group cooperation and coordination to achieve goals that could not be achieved—or at least could not be achieved as effectively—by individual means (Boyer, 2018; Tooby & Cosmides, 2010). Sometimes these goals are narrow and explicit. Because the cooperation that coalitions provide is so beneficial to group members, however, many coalitions operate as general-purpose *amplification coalitions*, amplifying the ability of their members to advance their individual interests

through open-ended mutual aid, social support, networks of trust, and opportunities for collective action (Tooby & Cosmides, 2010). Indeed, many coalitions that are ostensibly organised around narrow goals—for example, a religious community that congregates for collective worship—often function as amplification coalitions in practice, providing a much larger range of communal benefits for group members.

This dependence of coalitions on cooperation makes them highly vulnerable to free-riders, that is, individuals who reap the benefits of within-group cooperation without paying the costs required to sustain such benefits for others (Olson, 1965). If free-riding becomes practically advantageous, it will proliferate throughout a coalition and it will go extinct. Thus, a coalition's survival—not to mention its success in high-stakes competition with other coalitions of the sort that is characteristic of many social contexts—is dependent on its ability to solve this problem.

One solution—sometimes called a "control strategy" (Noe & Voelkl, 2013)—involves the enforcement of cooperative norms by means of surveillance and sanctions. Such methods are costly, however, and give rise to second-order collective action problems of their own. A different—albeit potentially complementary—"choice strategy" involves choosing group members from the outset in accordance with their *ingroup commitment*, where the goal "is not to force cooperation from those with whom you are fated to interact" but "to choose well in the first place, and to be chosen by others" (Sterelny, 2012, p. 102; see Noe & Voelkl, 2013).

Of course, an individual's degree of ingroup commitment is not an observable trait. Thus, coalitions that implement this strategy need to develop ways of distinguishing individuals by their degree of ingroup commitment. Given the benefits associated with the recognition of this trait, however, there are powerful incentives for exaggeration and outright deception. A large body of research in the psychological and social sciences draws on signalling theory to understand how groups solve this communication problem.

2.2 | Amplifying defection costs

Groups often require seemingly arbitrary costs, sacrifices, and restrictions on the part of their members that can seem puzzling. Given the assumption of rational self-interest, for example, one would expect groups that impose costs on group members to be less successful—because less attractive—than other groups. In fact, the opposite often appears to be true, at least up to certain limits (Iannaccone, 1992). According to an influential research tradition, many kinds of costs and sacrifices associated with group membership should be understood in terms of the role that they play in signalling and detecting signals of ingroup commitment (Bulbulia & Sosis, 2011; Iannaccone, 1992; Sterelny, 2020). The specific signalling strategy that I will focus on in this article involves interventions that signal commitment to one group by "amplifying defection costs" (Sterelny, 2020)—that is, by increasing the costs of, or in some cases eliminating the possibility of, defection to other groups. As Sterelny (2012, p. 296) puts it, "[b]y making migration between groups more difficult, such signals make individuals within groups more trustworthy".

Real-world examples of this signalling strategy appear to be ubiquitous. For example, group-specific body modifications—from gang tattoos in prisons to the radical forms of scarification practiced by many tribes—permanently identify the individual as a member of a group in a way that makes it difficult or even suicidal to try to join other groups (Boyer, 2018, Chap. 2). Similarly, it is plausible that group-specific traits, clothes, rituals, and restrictions that appear "funny" or "odd" to outsiders and thus invite stigma and ridicule are sometimes selected *because of this effect* (Iannaccone, 1992). Such strategic alienation from outsiders can take much

more extreme forms, however. For example, Sterelny (2012, pp. 295–296) notes that the initiation rituals of gangs, fringe subcultures, and cults often involve "high-price defection from the norms of larger society", which he illustrates with reference to a notorious New Zealand motorcycle gang in which prospective members must commit an act of rape as a condition of entry. Similar strategies are used by mafias around the world, where prospective members must often "make their bones" (i.e., commit a crime, often murder) to demonstrate commitment and gain membership (Dixit, Skeath & McAdams, 2021, p. 332).

The game-theoretic rationale underlying such diverse examples is similar. Coalitions can only identify genuinely committed group members if such members "can credibly commit to continued cooperation even if the opportunity for profitable defection arises" (Sterelny, 2012, p. 261). Acts that amplify defection costs—that is, that make it more costly and thus more difficult to join outgroups—constitute what Schelling (2006) calls an "external solution" to this commitment problem, whereby agents commit to a course of action by intervening in their environments—in this case, their social environments—to make it the rational course of action.

Of course, the proposal in this work is not that agents are consciously aware of the signalling function of such behaviours. For example, costly rituals, sacrifices, and restrictions in the context of religious groups are typically accompanied by bizarre rationales among group members (Iannaccone, 1992; Sterelny, 2020). A signalling theory of group behaviours claims that such intuitive explanations are mistaken. It thus purports to be a surprising theoretical discovery, uncovering an underlying rationale for otherwise puzzling group behaviours grounded in unconscious motivations to signal ingroup commitment and respond to the social rewards that attend such signals. Given the large body of independent evidence attesting to people's concern with ingroup acceptance and frequent ignorance of the motivations that underlie their actions, this perspective seems legitimate (see Boyer, 2018, Chapter 2; Pietraszewski, 2016).

2.3 | The strategic absurdity hypothesis

According to what I am calling the *strategic absurdity hypothesis*, the game-theoretic dynamics just described play a role in the genesis and maintenance of the absurd beliefs embraced within certain communities. Specifically, it proposes that it is precisely because such beliefs are viewed as absurd by outgroup members that their sincere and conspicuous endorsement credibly signals commitment to ingroup members.

Strategic absurdity hypothesis (SAH): The embrace and advertisement of absurd beliefs within certain coalitions (i) functions as a signal of ingroup commitment (ii) that is credible because the reputational damage incurred in the eyes of outsiders conspicuously amplifies defection costs.

As noted in the introduction, the SAH is increasingly influential, especially among those that take an evolutionary approach to understanding human psychology (Kurzban & Christner, 2011; Mercier, 2020; Petersen, 2020; Simler & Hanson, 2017; Williams, 2021). The earliest statement that I am aware of comes from Kurzban and Christner (2011), who argue that religious beliefs function as signals of ingroup commitment on the grounds that their striking irrationality makes the believer a less attractive recruit for competing groups. "Beliefs that preclude membership in other groups", they write, "are valuable because of commitment", and religious beliefs allegedly have this effect because they "make one appear, to those who do not share such beliefs, mentally ill" (Kurzban & Christner, 2011, p. 5). Simler and Hanson (2017, p. 279) endorse a similar idea in proposing that "the craziness of religious beliefs can function as a barometer of how strong the community is".

Although such proposals are in line with the SAH, it is misguided to apply this hypothesis specifically to religious beliefs. Some religious beliefs may not be irrational, many of those that are irrational no doubt have different underlying explanations (a point that I return to shortly), and many of the most plausible candidates for strategic absurdities—for example, extremely irrational conspiracy theories, ideological convictions, or nationalist myths—are not distinctively religious. Furthermore, it is too strong to say that absurd beliefs make one appear mentally ill to outsiders. Although this reaction might be generated by the beliefs embraced within certain extreme sects and cults, the reputational damage that results from being seen to hold absurd beliefs comes in degrees and is typically much weaker than this.

In recent work, Mercier (2020) extends the logic of the SAH to non-religious but still bizarre beliefs (see also Petersen, 2020; Petersen et al., 2020; Simler & Hanson, 2017; Williams, 2021). For example, focusing on absurd conspiracy theories such as "Pizzagate", which alleged that Hilary Clinton was running a child trafficking operation from the basement of a Washington pizzeria, Mercier (2020, p. 193) argues that the function of such bizarre beliefs is to burn the believer's bridges to the broader social world:

> By burning your bridges with as many of the competing groups as possible—making you unclubbable, as cognitive scientist Pascal Boyer put it—you credibly signal to the remaining groups that you'll be loyal to them, since you do not have any other options. (Mercier, 2020, p. 193).

Once again, however, the metaphor of burning bridges can be slightly misleading. The amplification of defection costs comes in degrees. Burning bridges—making oneself literally unclubbable—is best viewed as the extreme point of a continuum of signalling strategies. In general, the greater the conspicuous amplification of defection costs, the greater the degree of commitment that is signalled (Fessler & Quintelier, 2013; see Section 4.1). In many cases, the embrace of group-specific absurdities results in costly but not devastating reputational damage in the eyes of outgroup members. This is a small point, however, and the current article can be understood as clarifying the theoretical foundations of Mercier's proposal and exploring its empirical predictions.

The SAH also fits squarely within a more general research programme in philosophy and psychology focusing on the signalling or expressive functions of beliefs (see Funkhouser, 2017, 2020; McKay & Ross, 2020; Simler & Hanson, 2017; Williams, 2021). Specifically, strategic absurdities are social signals in the broad sense outlined by Funkhouser (2017): They are objects (i.e., beliefs) and the behaviours that such beliefs give rise to that are (1) formed and maintained in order to communicate information about the believer's ingroup commitment (2) to be detected by both ingroup and outgroup members (3) in order to modify their behaviour (i.e., to increase ingroup members' willingness to trust and cooperate with the believer and harm the believer's reputation in the eyes of outgroup members). Nevertheless, there are multiple ways that beliefs might perform signalling functions in the context of groups that are distinct from strategic absurdities. I will highlight four.

First, group attachments are often associated with pro-ingroup beliefs and anti-outgroup beliefs, including both direct evaluations (e.g., "Liverpool are the best", "Man United are rubbish") and their indirect consequences (e.g., "Liverpool was cheated out of the championship"). Such beliefs are often explained within social identity theory in terms of the regulation of "self-esteem", but their ultimate function might derive from signalling: A genuine and well-advertised belief in a group's superiority signals one's commitment to the group (Funkhouser, 2017, 2020). If so, this would constitute a different kind of signal to the one explored in this paper. Such beliefs signal ingroup commitment not by amplifying defection

costs but because of a reliable connection between their content—for example, believing that one's group is the best—and being committed to the relevant group.

Second, although some religious beliefs plausibly function as strategic absurdities, there are alternative means by which such beliefs might signal cooperative tendencies. One possible example is belief in powerful supernatural beings whose rewards and punishments track conformity to and deviation from group-specific norms (Bulbulia, 2004; Funkhouser, 2020). In this case, the beliefs might function as signals of ingroup trustworthiness, but they are fundamentally different from strategic absurdities. Once again, they would achieve their signalling function not through the amplification of defection costs but because of a reliable connection between their content—namely, belief in powerful moralising supernatural entities—and conformity to group-specific norms. That is, even if such beliefs are irrational, they are not functional because they are irrational.

Third, even though strategic absurdities involve costs, they follow a distinct logic to the concept of "costly signalling" as it is typically understood (see Smith & Harper, 2003). Costly signalling occurs when the honesty of a signal is guaranteed by its association with strategic costs that dishonest signallers are either unable or unwilling to pay.[4] It is an influential idea in the social sciences that many sacrifices, restrictions, and rituals associated with religious groups and other kinds of coalitions are costly signals of commitment to ingroup cooperation in this sense (Iannaccone, 1992). That is, because people would allegedly not incur such costs unless they were genuinely committed to the relevant group and its cooperative norms, the costs credibly signal such commitment. One might extend this idea to the beliefs that underlie such costly practices (Funkhouser, 2020). Importantly, the logic of strategic absurdities is subtly different. Acting—or believing—in ways that amplify defection costs does not merely *reveal* ingroup commitment but conspicuously and materially alters the individual's options and incentives in a way that makes sticking with the group the rational course of action. As Sterelny (2012) puts, such "signals constrain choice, and constrain it publicly, rather than reveal agent quality" (p. 111).

Finally, some have proposed that group beliefs can be biased towards irrationality merely through the role that they play in group identification (see Tooby, 2017). Consider an everyday mundane proposition—for example, that Paris is the capital of France. Given that people will believe this obvious truth no matter what groups they belong to, it would be useless as a badge of group identity. By contrast, nobody is likely to arrive at the kinds of strange and evidence-resistant religious or ideological convictions characteristic of many communities through a dispassionate search for truth. For this reason, they can function as useful markers of group membership. Once again, this rationale is fundamentally different from strategic absurdities, which are not mere badges of group identity but signals of ingroup commitment that are effective not because they are group-specific but because they damage the reputation of believers in the eyes of outsiders.

To summarise, then, strategic absurdities constitute a distinctive kind of signalling strategy in which individuals display their ingroup commitment by forming and displaying beliefs that result in selective reputational damage in the eyes of outgroup members. Nevertheless, it is important to stress that the foregoing signalling strategies are not mutually exclusive and likely interact, co-occur, and complement each other in particular cases. As I return to below (Section 4), for example, the logic of strategic absurdities is content-neutral in the sense that such beliefs can in principle concern anything, but it might be that beliefs selected to perform some of the signalling functions just outlined often generate the subject matter of such beliefs.

[4]More specifically, costly signalling involves differential costs, such that the costs of deception outweigh its benefits in a way that is reversed for honest signallers (see Smith & Harper, 2003).

Furthermore, even though strategic absurdities and costly signalling constitute logically distinct strategies, they likely complement each other. For example, incurring reputational damage in the eyes of outsiders plausibly does not only alter the individual's material incentives but also reveals something about the motivations of individuals willing to pay such a cost, and the general sacrifices often associated with holding such beliefs likely also convey important information about such motivations.

3 | THE PSYCHOLOGY OF STRATEGIC ABSURDITIES

How should the SAH be understood, psychologically? This question is usefully divided into two parts. First, what kind of mental states are strategic absurdities? Second, what kinds of psychological processes underlie the formation and maintenance of such mental states?

3.1 | Strategic absurdities as socially adaptive beliefs

First, it is useful to step back and consider how belief and belief formation should be understood more generally. When we believe a proposition *p*, we take it to be the case that *p*. Beyond this platitude, there is persistent controversy over how to understand the function(s) of beliefs, their defining characteristics, and the rationality of belief formation. According to one influential view in both philosophy and psychology, beliefs "aim at the truth" (Williams, 1973; see Fodor, 1975; Millikan, 1984). That is, they constitute the "maps by which we steer" (Ramsey, 1990, p. 146), functioning to provide us with accurate propositional information that we can exploit in guiding our inferences and decisions. Like maps, the utility of beliefs thus closely tracks their accuracy. Just as you will be more successful in navigating a territory with an accurate map, you will be more successful in achieving your goals—that is, in navigating reality more generally—if you have true beliefs.

Given this representational function, one would expect beliefs to have characteristics specialised for this function. Most obviously, if their function is to track how things are, they should be highly responsive to evidence which concerns how thing are. In addition, one would expect beliefs to be "inferentially promiscuous" (Stich, 1978): Insofar as beliefs function to provide accurate information about the world, agents should be eager to exploit such information in inference and decision-making whenever it is relevant. Finally, one would expect beliefs to be "practical-setting independent" (Van Leeuwen, 2014): Unlike psychological phenomena such as imaginings and suppositions, the information that beliefs provide should constitute the stable background model of reality that we draw upon across all practical contexts. According to many theorists, these expectations are vindicated when we turn to how belief formation actually works. That is, although it is widely acknowledged that there are various distorting factors that can interfere with such characteristics, many psychologists and philosophers hold that beliefs are in fact largely evidence responsive, inferentially promiscuous, and practical setting independent, at least when produced and maintained by properly functioning psychological mechanisms (Millikan, 1984; Stich, 1978; Williams, 1973).

There is a growing recognition that this conception of beliefs neglects the crucial social dimension of human belief formation (Boyer, 2018; Funkhouser, 2017; Mercier, 2020; Petersen et al., 2020; Simler & Hanson, 2017; von Hippel & Trivers, 2011; Williams, 2018, 2021). If our beliefs are the maps by which we steer, they are maps that are objects of intense social scrutiny.

We spend much of our time communicating and attributing beliefs, and we respond to people differently as a consequence of which beliefs we attribute to them in ways that can have dramatic effects on their wellbeing and social standing. Furthermore—and crucially—the kinds of beliefs that maximise social rewards are not identical with the kinds of beliefs that are best supported by one's available evidence. Thus, individuals who factor such social incentives into the process of belief formation would be at a competitive advantage relative to individuals who do not (Funkhouser, 2017; von Hippel & Trivers, 2011; Williams, 2021). For this reason, multiple theorists have argued that we should expect *socially adaptive beliefs*, that is, beliefs formed and maintained because of their effects on other people (see Williams, 2021 for a review).

Importantly, one would expect socially adaptive beliefs to have different characteristics to ordinary world-modelling beliefs. For example, because their function is not to track how things are, one would not expect them to be highly responsive to evidence concerning how things are (Funkhouser, 2017; Williams, 2021). Similarly, because socially adaptive beliefs are not truth-tracking, one would expect individuals to be reluctant to draw inferences from or to act on such beliefs when such inferences and actions are not themselves socially rewarded (Levy, 2018; Mercier, 2020). That is, one would expect such beliefs to be somewhat compartmentalised in the believer's mental life, or else individuals would risk letting inaccurate information compromise their general understanding of reality. Finally, one would expect individuals to be most likely to let such beliefs influence their thought and action in those practical settings in which the target audience of those beliefs exist. That is, one would expect them to be at least somewhat practical setting *dependent* (Levy, 2018; Van Leeuwen, 2014). Importantly, several philosophers have argued that when we turn to the most plausible candidates for socially adaptive belief—for example, ideological convictions, conspiracy theories, religious beliefs, and self-aggrandising beliefs—this is often the functional profile that we find: evidence-resistant beliefs partially cordoned off from the rest of the believer's mental life that seem to play a much more central role in guiding thought and action in some practical contexts than in others (see Levy, 2018; Mercier, 2020; Van Leeuwen, 2014; Williams, 2021).

Given such functional differences, it is tempting to infer that socially adaptive beliefs are either not genuine beliefs at all or constitute a distinct kind of belief. For example, perhaps socially adaptive beliefs should be assimilated to one of the psychological kinds introduced in the many attempts to draw distinctions among the class of cognitive attitudes, such as between beliefs and opinions (Dennett, 1978), beliefs and avowals (Rey, 1988), factual beliefs and religious credences (Van Leeuwen, 2014), intuitive and reflective beliefs (Sperber, 1997), full-fledged and fledging beliefs (Levy, 2018), and more.

I think that this inference should be resisted. The functional differences just mentioned are differences in degree. As I return to shortly (Section 3.2), socially adaptive beliefs generally—and strategic absurdities specifically—are not wholly decoupled from evidence and norms of epistemic rationality, and they do have downstream effects on the relevant believer's inferences and decision-making. This is because socially adaptive beliefs are dependent for their function on others thinking of them as genuine beliefs (Williams, 2021). This generates an important trade-off between the benefits of compartmentalising socially adaptive beliefs and the benefits of being attributed full commitment to such beliefs by others. That is, to the extent that it is clear to other agents that a belief is "accepted only nominally" (Mercier, 2020, p. xviii), those social benefits that accrue from conveying the believer's genuine commitment to the relevant belief will be diminished.

Consider strategic absurdities, for example. Although the mere avowal of absurd or outrageous claims can result in reputational damage, this damage will be much weaker if it is clear

that such statements are insincere or not genuinely believed, especially when it comes to evaluations of the individual's rationality and competence. As noted above, the greater the amplification of defection costs, the greater level of commitment that is signalled. Given this, one would expect strategic absurdities to more closely approximate the downstream properties of ordinary world-modelling beliefs as the degree of ingroup commitment that agents benefit from communicating increases (see Petersen et al., 2020, p. 23). That is, it is plausible that the degree of reputational damage a belief generates is dependent not just on its content but on the degree of conviction with which it is held, where the appearance of such conviction is inversely related to its compartmentalisation. In fact, as I return to below (Section 4.1), this inverse relationship between the degree of commitment being signalled and the degree of compartmentalisation is plausibly exactly what is found when it comes to absurd group beliefs. For example, the fact that a "totalist" group ideology has profound psychological and behavioural ramifications among adherents and cannot coexist with ordinary beliefs and activities constitutes a distinguishing mark of tight-knit cults and extremist groups in the sociological literature (Stein, 2017, p. 42). Thus, Mercier's (2020, p. 14) claim that "most popular misconceptions remain largely cut off from the rest of our minds and have few practical consequences" is potentially too strong. Such compartmentalisation applies to many socially adaptive beliefs, but there is also significant non-random variation in the degree of such compartmentalisation.

Stepping back, then, although it is correct to draw attention to the strange functional properties often exhibited by socially adaptive beliefs (e.g., Levy, 2018; Mercier, 2020; Williams, 2021; see Section 5), thinking in terms of distinct psychological kinds with fixed functional properties will blind us to this principled variation in the costly influences of such beliefs on the believer's thought and actions. With respect to strategic absurdities, for example, it is preferable to think of a hierarchy ranging from mere avowals through to beliefs with few behavioural ramifications through to beliefs that come to more closely approximate the downstream consequences of ordinary world-modelling beliefs, where such variation is driven by variation in the amplification of defection costs that individuals benefit from displaying.

3.2 | Strategic absurdities and motivated cognition

How should we understand the psychological processes that underlie the formation and maintenance of strategic absurdities when they go beyond mere avowals? In general, socially adaptive beliefs are best understood as resulting from independently well-established and more general tendencies to engage in motivated cognition, where "motivated cognition" describes the role of non-epistemic motivations in biasing individuals towards desired beliefs by unconsciously adjusting the way in which they seek out, neglect, interpret, and process information (see Kunda, 1990; Williams, 2021). In describing such tendencies as "more general" than socially adaptive belief formation, I mean that many examples of motivated cognition are driven by non-social motivations that would exist in the absence of other agents. Nevertheless, some forms of motivated cognition are forms of socially motivated cognition, in which the non-epistemic motivations that bias individual away from truth concern the believer's standing in the social world and give rise to beliefs adapted to it (see von Hippel & Trivers, 2011; Williams, 2020).

Although there is much that we do not understand about motivated cognition, the psychological literature does provide several clear lessons that can be used to understand and illuminate the SAH (see Kunda, 1990; Westen, Blagov, Harenski, Kilts & Hamann, 2006). First,

motivated cognition is goal driven. These goals need not be—and often will not be—conscious, but in order to underpin motivated cognition they must come into conflict with the acquisition of knowledge and thus motivate individuals to treat information in biased ways (Kunda, 1990). In the case of strategic absurdities, the relevant goal is to advertise the believer's ingroup commitment. If individuals come to recognise—even if only implicitly—that embracing and advertising certain unfounded beliefs would advance this goal, it can come into conflict with the search for truth and thus motivate a biased search and interpretation of evidence.

Second, motivated cognition is emotion laden. In general, the beliefs that we hold powerfully influence how we feel, both because of the states of affairs that they represent but also because of how they contribute to or hinder our goals. For example, if I desire to think well of myself and come across evidence that I have acted badly, I may be pushed towards a belief that I find distressing. The drive to alleviate this aversive emotional experience might then motivate me to wilfully ignore and misinterpret evidence concerning my transgression. In this way motivated cognition typically involves "a form of implicit emotion regulation in which the brain converges on judgments that minimize negative and maximize positive affect states associated with threat to or attainment of motives" (Westen et al., 2006, p. 1947), where the kinds of affective—or, more generally, emotional—states being regulated are specific to the motives involved. If the SAH is correct, for example, one would expect strategic absurdities to involve the distinctive family of social emotions connected to the motivational systems underlying group affiliation and cooperation, where success results in feelings of security, belonging, and pride, and failure results in aversive feelings of anxiety, loneliness, and shame (Del Giudice, 2018, pp. 36–45; see Section 4.1). As with displays of ingroup commitment more generally, however, such emotion regulation is merely the proximate mechanism underlying the ultimate function of ingroup signalling (see Boyer, 2018, p. 50).

Third, motivated cognition is indirect. That is, individuals do not arrive at desired beliefs directly through intentional choice but indirectly "through reliance on a biased set of cognitive processes—that is, strategies for accessing, constructing, and evaluating beliefs" (Kunda, 1990, p. 480). For example, when we engage in motivated cognition, we will often seek out evidence that confirms the desired belief and ignore or misinterpret evidence in tension with it, apply highly asymmetric evidential and justificatory standards to any evidence and arguments that we do encounter, opportunistically assign and withhold trust to testimony in accordance with its congruence or incongruence with the desired belief, and so on (see Funkhouser, 2019; Kahan, 2013; Kunda, 1990). In this way motivated cognition influences information sampling and processing in ways that go far beyond biased conscious reasoning.

Finally, and relatedly, motivated cognition is subject to a rationalisation constraint. That is, even though it involves forming and maintaining a belief for reasons independent of its truth, we can only bring ourselves to believe things for which we can find genuinely epistemic justifications, even if such justifications constitute post hoc rationalisations (Kunda, 1990). This can create the illusion that the *cause* of the relevant belief is identified by the reasons provided in its rationalisation, akin to concluding that the cause of costly rituals is provided by the intuitive but implausible justifications people often provide for such behaviours (see Sterelny, 2020). Thus, even when it comes to genuinely absurd—that is, extremely irrational—beliefs, one would still expect such beliefs to be associated with a kind of pseudo-epistemic support. This drive to identify epistemic justifications distinguishes socially adaptive beliefs from other non-truth tracking mental states such as imaginings and suppositions and imposes at least a minimal constraint on the irrationality of such beliefs. As demonstrated by the feeble nature of the justifications provided for many outlandish beliefs, however—for example, those who point to

testimony from anonymous users on message boards and hidden meanings and code words in Trump speeches as justifications of QAnon[5]—this constraint is extremely weak and consistent with striking forms of epistemic irrationality.

Stepping back, then, the dual concepts of socially adaptive belief and motivated cognition provide a framework for extending the signalling framework for understanding puzzling group behaviours and observable traits outlined in Section 2 to the formation and maintenance of puzzling group beliefs. That is, strategic absurdities—if they exist at all—should be understood as an example of a more general phenomenon of socially adaptive beliefs grounded in independently well-established motivational biases when it comes to sampling and processing information.

4 | IS THE STRATEGIC ABSURDITY HYPOTHESIS PLAUSIBLE?

So far, I have described nothing more than a schematic "how possibly" model. That is, I have not provided any reasons to *endorse* the SAH as an explanation of a specific class of absurd group beliefs. My chief aim in this article is to clarify the theoretical foundations of this hypothesis. Evaluating its plausibility when it comes to explaining the origins and maintenance of specific group beliefs is a task for future work. Nevertheless, as an initial contribution to this project, I will conclude by identifying several predictions that the SAH as described in the foregoing section generates. Before this, however, two important qualifications are necessary.

First, the SAH is obviously not intended to explain all cases in which groups harbour seemingly irrational beliefs. In some cases, for example, group-specific beliefs that seem irrational to outsiders might be a reasonable response to misleading evidence, testimony, and arguments. In other cases, group-specific beliefs that seem irrational are irrational, but this irrationality is driven by other motivations and biases, including different signalling strategies of the kind described in Section 2.3 (see also Funkhouser, 2020). Nevertheless, the SAH is especially well-suited to explaining cases in which groups organise around *extremely* irrational beliefs (see Mercier, 2020, pp. 194–195). This is ultimately an empirical conjecture that must be evaluated on a case-by-case basis, however.

Second, even if the SAH does correctly describe the psychological and social underpinnings of some specific class of group beliefs, it is incomplete as an explanation. As noted above (Section 2.3), it abstracts away from—and thus itself offers no explanation of—the contents of absurd beliefs in specific cases. In this way the SAH identifies a kind of doxastic strategy that might be applied across a wide variety of different (e.g., religious, political, cultural, etc.) contexts. That is, it purports to explain why a group would sometimes gravitate towards group-specific absurd beliefs, but on its own it does not explain why any given group would gravitate towards any specific absurd beliefs.

In general, this question of content determination for strategic absurdities will likely not have a unified answer but will be dependent on a range of context-specific psychological, social, institutional, and historical factors. Nevertheless, there are several kinds of themes that are likely to recur when it comes to strategic absurdities. As noted above (Section 2.3), groups often form group-serving and group-aggrandising beliefs, anti-outgroup beliefs, and religious or quasi-religious beliefs for independent reasons, and such beliefs might often provide the

[5]See Doward (2020).

contents or subject matter for strategic absurdities. Furthermore, certain kinds of belief contents are likely especially well-suited to selectively damaging the reputation of ingroup members in the eyes of outgroup members, including bizarre supernatural themes and beliefs that demonise outgroup members in outrageous ways (Mercier, 2020, pp. 194–196). Exploring this important issue concerning content is crucial for understanding the SAH and its application to specific cases, but it is a task that falls beyond the scope of this article.

Despite its incomplete nature, the SAH does generate certain predictions. Although these predictions are not individually diagnostic, in conjunction with each other they would present a compelling case that the SAH should at least be taken seriously as a candidate explanation for a specific class of beliefs. Given space constraints, I will simply identify the predictions here, and leave the additional—and difficult—issue of exploring how experimental or observational data could be brought to bear on such predictions to future work.

4.1 | Predictions of the strategic absurdity hypothesis

4.1.1 | Ingroup commitment and outgroup stigma

Most obviously, the SAH predicts that the beliefs that it applies to will in fact be associated with the relevant pattern of ingroup commitment and outgroup stigma that it describes. That is, sincere endorsement of the relevant group beliefs should be treated as signals of ingroup commitment by fellow group members and result in some kind of reputational cost in the eyes of outgroup members. Of course, the fact that some set of group beliefs elicits such responses does not constitute strong evidence for the SAH. The pattern may be causally irrelevant to the formation of such beliefs, for example. Nevertheless, a failure to observe this pattern of responses would constitute decisive evidence against the SAH, because such responses constitute an essential part of the alleged mechanism by which belief formation is supposed to work on this hypothesis.

4.1.2 | Signalling without persuasion

The SAH also predicts that individuals should be motivated to display their adherence to the relevant beliefs. That is, insofar as the—or at least a—function of such beliefs is to advertise the believer's ingroup commitment, sincere belief must be advertised. On the surface, this eagerness to display the beliefs is difficult to make sense of on non-signalling explanations (see Funkhouser, 2020; Williams, 2020). That is, if individuals form the relevant beliefs not to signal their commitment to a group but for purely epistemic reasons, why should they go to great lengths to display their commitment to this belief to others?

This is too quick, however. If I believe that the Democratic Party is run by Satan-worshipping cannibalistic paedophiles, I might have several reasons for wanting to communicate this belief. For example, I will presumably be eager to persuade at least some other agents of its truth. Thus, the mere fact that individuals are eager to communicate the contents of a belief is not strong diagnostic evidence for a social signalling explanation. Nevertheless, signalling explanations of beliefs do predict a distinct kind of communication, which I will call *signalling without persuasion*: If individuals form a belief in order to signal their ingroup commitment, they should be highly motivated to advertise their endorsement of this belief in ways that cannot be explained by the intention to persuade anyone of its truth. Indeed, insofar as such beliefs are selected and displayed to elicit

reputational damage in the eyes of outgroup members, one should expect them to be communicated in ways optimised for *not* persuading such agents of their truth.

Although I am not aware of any systematic empirical research that bears on this prediction, signalling without persuasion seems to be common when it comes to irrational group beliefs. For example, commitment to such beliefs is often displayed with jewellery, tattoos, bumper stickers, and—in the case of members of the flat Earth society—T-shirts.[6] Furthermore, whereas genuine persuasion is dependent on strong arguments, cues of benevolence, and common interest (Mercier, 2020), absurd group beliefs are often rationalised with a kind of pseudo-epistemic support that nobody outside of the community finds remotely compelling—recall the hidden code words in Trump's speeches as evidence for QAnon—and communicated in ways that actively demonise non-believers as stupid or evil (Funkhouser, 2020, p. 7; Mercier, 2020, p. 195). Such communicative strategies are difficult to make sense of under the assumption that such beliefs constitute genuine attempts to understand reality, but they are straightforwardly predicted by the SAH.

4.1.3 | Extreme irrationality

A third prediction that follows directly from the SAH is that beliefs alleged to signal ingroup commitment should in fact be extremely irrational. That is, individuals should treat such beliefs—and evidence and reasons which bear on their truth—in highly biased ways. Of course, the rationalisation constraint introduced above places some constraints on what individuals can bring themselves to believe, but such constraints are extremely weak. For example, Doomsday cults whose members discover that the end of the world does not arrive on the day that they predicted it would often *increase* their confidence in their cults' beliefs upon encountering the decisive disconfirmation of their prediction (Dawson, 1999). Likewise, those who endorsed Pizzagate revealed no evidence of reducing their confidence in the conspiracy after an armed man broke into the relevant Washington pizza parlour to discover a complete absence of any predicted evidence (Doward, 2020). In such cases, the adherents of the relevant beliefs develop justifications for disregarding these striking falsifications, but it is difficult to deny that this response constitutes a profound deviation from ordinary—not to mention optimal—inferential practice.

4.1.4 | Non-epistemic dynamics

The SAH predicts that absurd group beliefs should not be undermined by contrary evidence. Such beliefs do change, however, and new ones come into existence. The SAH predicts that such dynamics are driven not by epistemic considerations but by the non-epistemic dynamics of variations in ingroup and outgroup response highlighted in the first prediction. Thus, not only does the SAH predict that absurd group beliefs should be almost completely impervious to contrary evidence and reasons, but it also predicts that it should be possible to manipulate people's attitudes towards such beliefs by means of a distinctive kind of information that is *irrelevant to their truth*: namely, information about how ingroup members and outgroup members respond to such beliefs.[7] Of course, it is difficult to disentangle these two potential causes of

[6]As shown in the documentary *Behind the curve*.

[7]See Tooby and Cosmides (2010) for an important discussion of how non-epistemic social factors and not truth content or informativeness determine the significance assigned to certain pieces of information.

belief change in practice. For example, a change in how fellow group members respond to a belief could be interpreted as evidence that bears indirectly on its truth through mechanisms of trust. Nevertheless, there is a principled distinction between epistemic and non-epistemic factors, and the SAH makes the clear prediction that it should be possible to adjust people's attitude towards strategic absurdities merely by manipulating non-epistemic factors about ingroup and outgroup responses.

4.1.5 | Content-independent social emotions

If I am right in connecting strategic absurdities to motivated cognition grounded in motivational systems pertaining to group affiliation and cooperation, one would expect to observe a distinct pattern of emotional states associated with the endorsement and communication of such beliefs. For example, such beliefs should be associated with a sense of group-based pride and individuals should treat evidence in tension with them not as useful information with which to update their beliefs but as a threat to their identity.

Crucially, the prediction here is not merely that absurd group beliefs are associated with powerful emotions. If I believe that NASA is lying to the world's population about the shape of the planet, I am unlikely to be dispassionate about this. To the extent that this belief functions as an ordinary world-modelling belief, however, my emotions will be connected to its *content*—that is, to the state of affairs that it represents. By contrast, if the SAH provides a correct explanation of the formation and maintenance of a given belief, it should be associated with a distinct family of content-independent social emotions involved in the general communication of ingroup commitment. For example, I might be *proud* to be the sort of person who believes that NASA is involved in hiding the shape of the planet, and I might react to evidence and arguments against this belief with the same kind of anxiety that attends identity threats more generally. Furthermore, the SAH predicts that this pattern of affective and emotional response should be found among absurd group beliefs with otherwise radically different contents (e.g., religious, ideological, conspiratorial, etc.) and also with public symbols and behaviours intended to advertise ingroup commitment.

4.1.6 | Compartmentalisation

A sixth important prediction of the SAH is that strategic absurdities should be functionally peculiar not just in terms of how they respond to evidence which bears on their truth but in terms of their downstream consequences on the believer's thoughts and actions. That is, insofar as group-specific absurdities are embraced not to track how things are but to advertise the believer's ingroup commitment, there should be an implicit recognition of this function and thus a reluctance to let such beliefs freely guide the believer's inferences and decisions. Furthermore, we should also expect this insulation of the belief from the rest of the believer's mental life and action to vary as a function of the degree of ingroup commitment that they benefit from communicating (see Section 3.1).

As noted above (Section 3), this partial compartmentalisation of belief is exactly the functional profile often found when it comes to many religious beliefs, evidence-resistant ideological convictions, and bizarre conspiracy theories (see Levy, 2018; Mercier, 2020). As Mercier (2020, p. 155) notes concerning Pizzagate, for example, of the millions of people who professed to endorse this

conspiracy theory—and appeared to be *sincere* in their avowal of it—only one person actually went to release the children allegedly being held captive in the pizzeria's basement. For the others, the principal way in which their conviction manifested itself in their behaviour was in their continual advertisements of the belief. A similar observation is often made about religious beliefs. For example, such beliefs seem at least somewhat practical setting dependent, exerting a much greater control over behaviour when made salient to people (Christian beliefs in the US notoriously influence Christian behaviour much more on Sundays than on other days), and they are famously also often encapsulated from the rest of the believer's network of beliefs and intuitive inferences (see Levy, 2018; Van Leeuwen, 2014). Such compartmentalisation is difficult to understand on non-signalling explanations of such beliefs, but neatly predicted by a social signalling explanation such as the SAH (see Levy, 2018; Mercier, 2020).

As also noted above (Section 3.1), however, there is important variation in the degree to which absurd beliefs influence costly behaviours outside of the context of explicit avowals. In many contexts, for example, absurd beliefs are not neatly insulated from the rest of the believer's mental life and behaviours. For example, religious beliefs are often associated with powerful sacrifices and restrictions, conspiracy beliefs drive people to reject life-saving vaccinations and storm government buildings, extreme ideological convictions motivate people to leave their families and fight wars, and so on. If the SAH as I have described it is correct, this variation should be non-random and linked to variation in the degree of ingroup commitment that agents benefit from signalling (see Petersen et al., 2020, p. 23). Although I am unaware of any evidence which bears directly on this prediction, it is notable that the distinction between churches and sects in the sociology of religion tracks stable behavioural differences between such groups that seem to be driven in large part by variation in the costs associated with signals of ingroup commitment (Iannaccone, 1992). Furthermore, as noted above (Section 3.1), there appears to be significantly less compartmentalisation when it comes to the "totalist" belief complexes found among extremely tight-knit cults and extremist groups, which are marked out in the sociological literature by their profound ramifications for their adherents' thoughts and actions (Stein, 2017, p. 42).

4.1.7 | Individual variability

Finally, the SAH generates distinctive predictions about individual variation in the willingness to embrace strategic absurdities. Such beliefs constitute a socially costly means of gaining access to an ingroup, one which involves acting in ways that harm the believer's reputation in the eyes of all outgroup members. Thus, one would expect variation in the execution of this strategy to vary as a function of its relative costs and benefits for different individuals. For example, for people with a high social standing with membership in lots of cross-cutting groups, the costs of such a strategy will be substantial. As a consequence, they will be less likely to embrace strategic absurdities, or at least—as with many mainstream religious communities in society—more likely to embrace such beliefs in ways that are sufficiently and conspicuously compartmentalised that the reputational damage that they generate is reduced (see Section 3.1 above). By contrast, ostracised individuals with a low social standing and weak social ties should be correspondingly more likely to execute costlier strategies for gaining access to amplification coalitions and the communal benefits that they provide.

There is some observational data that bears suggestively on this prediction. For example, ostracism is notoriously associated with a greater willingness to join extremist groups (Hales &

Williams, 2018), and a tendency towards extreme conspiracy theories is positively correlated with a perceived low social standing and weak social networks (Freeman & Bentall, 2017). However, such evidence cannot distinguish between whether the embrace of extreme beliefs leads to ostracism and social exclusion or whether—as predicted by the SAH—ostracism increases an individual's willingness to embrace group-specific extreme beliefs. The only experimental data that I am aware of that bears on this question is recent research demonstrating that inducing a sense of ostracism in experimental subjects increased their willingness to endorse unambiguously false claims associated with the political party to which they belong (Garrett, Sude & Riva, 2020). The authors use this finding to suggest that "citizens with the least social or economic power may be uniquely prone to seek social inclusion through the endorsement of unambiguously false claims made by members of their own political party" (Garrett et al., 2020, p. 168), a conclusion that is in line with the SAH. This is clearly a topic that demands significantly more research, however.

5 | CONCLUSION

The claim that I have tried to explore in this article is simple: For some communities, the embrace of group-specific absurdities functions as a display of ingroup commitment that is credible precisely because such beliefs are viewed as absurd by outgroup members. I have argued that this signalling strategy is best understood in terms of independently well-established mechanisms of motivated cognition, and I have sought to clarify how we should understand the kinds of mental states it involves and what distinctive predictions this proposal generates. Of course, a proper understanding and evaluation of the SAH requires that many more psychological and social details be filled in. When it comes to many of the predictions just outlined, for example, there is little or no empirical research addressing them, in significant part because the SAH has received little systematic attention in psychology and the social sciences. Nevertheless, I hope that this article contributes to motivating others to take up this and related tasks. If the SAH is correct, it has profound consequences for our understanding of ourselves and the social worlds that we inhabit, and important lessons for how to address the problems generated by widespread irrational beliefs in society. Although much work remains to be done, I hope that I have done enough to show that the SAH is not itself absurd, and that this work is worth pursuing.

ACKNOWLEDGEMENT

For helpful comments, discussion, feedback, and insights, I would like to thank two anonymous reviewers, Marcella Montagnese, Stephen Gadsby, Eric Funkhouser, Hugo Mercier, Sam Wilkinson, Vaughan Bell, Adrian Currie, and all of those who attended the Egenis seminar at the University of Exeter where I presented this work.

DATA AVAILABILITY STATEMENT

Not applicable.

ORCID

Daniel Williams https://orcid.org/0000-0002-9774-2910

REFERENCES

Boyer, P. (2018). *Minds make societies: How cognition explains the world humans create*. London: Yale University Press.

Bulbulia, J. (2004). Religious costs as adaptations that signal altruistic intention. *Evolution and Cognition*, *10*(1), 19–38.

Bulbulia, J., & Sosis, R. (2011). Signalling theory and the evolution of religious cooperation. *Religion*, *41*(3), 363–338.

Dawson, L. (1999). When prophecy fails and faith persists: A theoretical overview. *Nova Religio*, *3*(1), 60–82.

Del Giudice, M. (2018). *Evolutionary psychopathology: A unified approach*. New York, NY: Oxford University Press.

Dennett, D. (1978). How to change your mind. In D. Dennett (Ed.), *Brainstorms: Philosophical essays on mind and psychology* (pp. 300–309). Cambridge: MIT Press.

Dixit, A., Skeath, S. & McAdams, D. (2021). *Games of strategy* (5th ed.). New York, NY: W. W. Norton & Company.

Doward, J. (2020). .December 13, 2020. Retrieved from https://www.theguardian.com/world/2020/sep/20/the-qanon-conspiracy.

Fessler, D. M. & Quintelier, K. (2013). Suicide bombers, weddings, and prison tattoos. In K. Sterelny, R. Joyce, B. Calcott & B. Fraser (Eds.), *Cooperation and its evolution* (pp. 459–483). London: The MIT Press.

Fodor, J. (1975). *The language of thought*. New York, NY: Thomas Y. Crowell.

Freeman, D. & Bentall, R. P. (2017). The concomitants of conspiracy concerns. *Social Psychiatry and Psychiatric Epidemiology*, *52*(5), 595–604.

Funkhouser, E. (2017). Beliefs as signals: A new function for belief. *Philosophical Psychology*, *30*(6), 809–831.

Funkhouser, E. (2019). *Self-deception*. London: Routledge.

Funkhouser, E. (2020). A tribal mind: Beliefs that signal group identity or commitment. *Mind & Language*, 1–22. https://doi.org/10.1111/mila.12326

Garrett, R. K., Sude, D. & Riva, P. (2020). Toeing the party lie: Ostracism promotes endorsement of partisan election falsehoods. *Political Communication*, *37*(2), 157–172.

Hales, A. H. & Williams, K. D. (2018). Marginalized individuals and extremism: The role of ostracism in openness to extreme groups. *Journal of Social Issues*, *74*(1), 75–92.

Iannaccone, L. R. (1992). Sacrifice and stigma: Reducing free-riding in cults, communes, and other collectives. *Journal of Political Economy*, *100*(2), 271–291.

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*, 407–424.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498.

Kurzban, R. & Christner, J. (2011). Are supernatural beliefs commitment devices for intergroup conflict. In J. P. Forgas, A. Kruglanski & K. D. Willimas (Eds.), *The psychology of social conflict and aggression* (Vol. 13, pp. 285–300). Sydney Symposium of Social Psychology). New York, NY: Taylor and Francis.

Levy, N. (2018). Showing our seams: A reply to Eric Funkhouser. *Philosophical Psychology*, *31*(7), 991–1006.

McKay, R. T. & Ross, R. M. (2020). Religion and delusion. *Current Opinion in Psychology*, *40*, 160–166.

Mercier, H. (2020). *Not born yesterday: The science of who we trust and what we believe*. London: Princeton University Press.

Millikan, R. (1984). *Language, thought and other biological categories*. Cambridge: MIT Press.

Noe, R. & Voelkl, B. (2013). Cooperation and biological markets: The power of partner choice. In K. Sterelny, R. Joyce, B. Calcott & B. Fraser (Eds.), *Cooperation and its evolution* (pp. 131–151). London: The MIT Press.

Olson, M. (1965). *The logic of collective action: Public goods and the theory of groups*. Cambridge, MA: Harvard University Press.

Petersen, M. B. (2020). The evolutionary psychology of mass mobilization: How disinformation and demagogues coordinate rather than manipulate. *Current Opinion in Psychology*, *35*, 71–75.

Petersen, M. B., Osmundsen, M. & Tooby, J. (2020). The evolutionary psychology of conflict and the functions of falsehood. Working Paper.

Pietraszewski, D. (2016). How the mind sees coalitional and group conflict: The evolutionary invariances of n-person conflict dynamics. *Evolution and Human Behavior*, *37*(6), 470–480.

Ramsey, F. P. (1990). In D. H. Mellor (Ed.), *Philosophical papers*. New York, NY: Cambridge University Press.

Rey, G. (1988). Toward a computational account of akrasia and self-deception. In B. P. McLaughlin & A. O. Rorty (Eds.), *Perspectives on self-deception* (Vol. 6, pp. 264–296). Los Angeles: University of California Press.

Richerson, P. J. & Boyd, R. (2008). *Not by genes alone: How culture transformed human evolution*. Chicago, IL: University of Chicago press.

Schelling, T. (2006). *Strategies of commitment and other essays*. Cambridge, MA: Harvard University Press.

Simler, K. & Hanson, R. (2017). *The elephant in the brain*. Oxford: Oxford University Press.

Smith, J. M. & Harper, D. (2003). *Animal signals*. Oxford: Oxford University Press.

Sperber, D. (1985). Anthropology and psychology: Towards an epidemiology of representations. *Man*, *20*, 73–89.

Sperber, D. (1997). Intuitive and reflective beliefs. *Mind & Language*, *12*(1), 67–83.

Sterelny, K. (2012). *The evolved apprentice*. Cambridge: MIT press.

Sterelny, K. (2020). Religion: Costs, signals, and the Neolithic transition. *Religion, Brain & Behavior*, *10*(3), 303–320.

Stein, A. (2017). *Terror, love and brainwashing: Attachment in cults and totalitarian systems*. London: Routledge.

Stich, S. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, *45*, 499–518.

Tooby, J. (2017). Coalitional instincts. In *Edge*. Edge Foundation. Retrieved from https://www.edge.org/response-detail/27168

Tooby, J. & Cosmides, L. (2010). Groups in mind: The coalitional roots of war and morality. In *Human morality and sociality: Evolutionary and comparative perspectives* (pp. 91–234). London: Palgrave-Macmillan.

Van Leeuwen, N. (2014). Religious credence in not factual belief. *Cognition*, *133*, 698–715.

von Hippel, W. & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, *34*(1), 1–16.

Westen, D., Blagov, P. S., Harenski, K., Kilts, C. & Hamann, S. (2006). Neural bases of motivated reasoning: An fMRI study of emotional constraints on partisan political judgment in the 2004 US presidential election. *Journal of Cognitive Neuroscience*, *18*(11), 1947–1958.

Williams, B. (1973). Deciding to believe. In B. Williams (Ed.), *Problems of the self* (pp. 136–151). Cambridge, MA: Cambridge University Press.

Williams, D. (2018). Hierarchical Bayesian models of delusion. *Consciousness and Cognition*, *61*, 129–147.

Williams, D. (2021). Socially adaptive belief. *Mind & Language*, *36*(3), 333–354. https://doi.org/10.1111/mila.12294

Williams, D. (2020). Motivated ignorance, rationality, and democratic politics. *Synthese*, *198*, 1–21.

How to cite this article: Williams, D. (2021). Signalling, commitment, and strategic absurdities. *Mind & Language*, 1–19. https://doi.org/10.1111/mila.12392